

AD-A117 288

ARMY AVIONICS RESEARCH AND DEVELOPMENT ACTIVITY FORT--ETC F/6 17/2
VOICE INTERACTIVE SYSTEMS TECHNOLOGY AVIONICS (VISTA) PROGRAM.(U)
JUN 82 L W REED

UNCLASSIFIED

NL

2
■



ENT
DA
FEB 82
DTIC

CONT

A
728

①

18 JUN 1982

REED

AD A 11 7288

VOICE INTERACTIVE SYSTEMS TECHNOLOGY AVIONICS (VISTA) PROGRAM (U)

LOCKWOOD W. REED
US ARMY AVIONICS RESEARCH AND DEVELOPMENT ACTIVITY
FORT MONMOUTH, N.J. 07703

INTRODUCTION

As the complexity of aircraft missions increases, there is a commensurate increase in the number of functions and operations to be performed by the pilot and crew of a given aircraft. Recently, military single-seat rotary aircraft have been proposed, compounding the problem of the man-machine interface.

This problem can be significantly reduced if the following three basic requirements can be satisfied: (1) the pilot can maintain hands on flight controls throughout the entire operation of the aircraft (2) the pilot's visual attention can be directed toward the flying of the aircraft (particularly important for Nap-of-the-Earth flight) (3) the pilot can still control all necessary aircraft subsystems under the conditions of (1) and (2).

Preliminary investigations by the Avionics Research and Development Activity (AVRADA) and many other governmental and industrial concerns have indicated that an integrated system of voice recognition and voice response can meet all the above requirements.

In order to investigate the potential advantages of Voice Technology, AVRADA has initiated a program entitled Voice Interactive Systems Technology Avionics (VISTA). The VISTA program is taking a phased approach to the introduction of voice recognition and response equipment into Army aircraft. Before detailing the phases of the VISTA program it would be appropriate to describe here the AVRADA facilities which will support the VISTA program. AVRADA maintains an extensive Computer aided Design and Audio Analysis Laboratory. The Audio Analysis Laboratory consists of two sound chambers (one Anechoic Fig 1, and one Sound Absorption Fig 2).

DTIC FILE COPY

DTIC
ELECTE
S JUL 22 1982 D
B

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

82 07 19 227

REED

Each chamber is tied via audio signal and digital data lines into the test equipment rack (Fig 3) which contains signal measuring amplifiers, two audio spectrum analyzers, various filters and equalizers, and audio recording equipment. The equipment in the test rack is in turn connected to the main laboratory computer system (Fig 4) consisting of a 16-bit mini-computer, 17 terminals, line printer, parallel I/O to test equipment and 80 mbyte disk drive, and a high speed array processor. AVRADA's Audio Analysis laboratory contains one of the most extensive libraries of Army aircraft noise environment tape recordings. The recordings were taken in the field at various Army installations using AVRADA's portable audio analysis equipment which consists of a precision portable audio recorder, sound level meter with 1/3 octave and narrowband filter sets and strip chart recorder. Recordings were made in several OH-58, UH-1, AH-1, CH-47, OV-1D, CH-54, and a UH-60 aircraft in various modes of flight from hover to tactical maneuvers.

VISTA PROGRAM

The first phase of the VISTA program (Fig 5) consists of writing software for our in-house computer systems to aid in the testing and evaluation of voice recognizers (the software will be discussed in greater detail under test procedures), and the evaluation of selected off-the-shelf voice recognizers. The use of off-the-shelf recognizers has a twofold benefit. First, testing procedures applicable to all recognizers can be developed economically, and secondly, the performance of off-the-shelf recognizers can yield baseline cost/performance information to which more sophisticated recognizers can be compared. Recognizer testing during phase one will be performed exclusively in the Audio Analysis Laboratory. Testing in the aircraft will be limited to the generation of recognizer training/test tape recordings made in the aircraft to be used as a validation check on the accuracy of the chamber tests.

As the recognizer test/evaluation software and procedures are finalized, phase II of the VISTA program will begin (Fig 6). Phase II involves applying the testing software and procedures developed during Phase I to more sophisticated (and more costly) commercial and non-commercial voice recognizers. In addition, since high ambient noise is a prime concern for any attempted installation of voice recognition equipment in Army aircraft, testing will be performed by front-ending the above recognizers with off-the-shelf noise reduction devices in addition to the various noise canceling microphones already in use.



Unannounced Justification <input checked="checked" type="checkbox"/>	
By _____	
Distribution/ _____	
Availability Codes	
Dist	Avail and/or Special
A	

REED

As the predictability of various voice recognizers is established, Phase III (Fig 7) of the VISTA program will begin with the installation of selected recognizers in AVRADA's Systems Testbed for Avionics Research (STAR) aircraft. The STAR aircraft (Fig 8) consists of a UH-60 helicopter modified to include a MIL-STD-1553 data bus, and various avionic subsystems (i.e., radio, navigation, night pilotage, etc.), all of which are connected to the common 1553 data bus. Sound recordings and measurements have been made in this aircraft and are being used in chamber environment simulations of the aircraft. Figure 9 shows the acoustic frequency spectrum taken in the aircraft at two locations, the cockpit and midship. The sound level has been measured at 103 db on the A weighted scale in the cockpit and 107 dbA amidships. Testing in the STAR aircraft will primarily be directed toward applications and operations for voice recognizers in the aircraft. To obtain meaningful results from the testing it will be necessary for the recognition equipment to be integrated into the aircraft data bus and to have access to all subsystems. To accomplish this and still have the flexibility of evaluating many different types of recognizers, a problem is created since most non-commercial and all commercially available recognizers do not have any MIL-STD-1553 data bus interface.

The approach taken to solve the problem is to install a general-purpose militarized computer (Fig 10) which will serve many functions. First, the computer will act as an intelligent interface providing the MIL-STD-1553 bus interface hardware and the data bus control software for the aircraft side of the integration. Second, since the large majority of commercial and non-commercial recognizers communicate via RS-232 or RS-422 hardware, the computer will provide several of these interfaces. Third, the operating system software of the computer will be designed so as to minimize the impact of changing from one recognizer to another. Fourth, the computer will contain all the software for a given control scenario. As an illustration, the use of a recognizer to control the onboard radios will be totally controlled by the computer. At any point in the scenario the computer will restrict the recognizer to a specific subset of its vocabulary to reduce the chance of false recognitions. When an utterance is made and the recognizer outputs its best guess, the computer will decide what to do with the response, what equipment is to be affected, and what 1553 data bus message will be sent. By dividing the control scenario into specific software modules keyed to generic recognizer responses, minimal impact on the computer software is achieved when changing voice recognizers. Only that software which extracts the generic information from a particular recognizer message need be changed.

To complement the contribution of the industrially available recognizers to the VISTA program, work will begin in house during Phase IV (Fig 11) to develop and test voice recognition and noise cancellation algorithms

REED

tailored to the Army aircraft environment. The high speed array processor mentioned previously will be used exclusively for voice recognition and noise cancellation algorithm testing, to permit real time response.

The emphasis of the VISTA program has been directed toward the voice recognition problem because technically it is the more risky half of an integrated voice interactive system. However, voice response does present unique problems of its own primarily in the area of human factors. Technically there are many implementations of voice response available, each offering certain advantages. For the most part, the selection of a given voice response unit is a tradeoff between intelligibility and digital storage capacity but even here the technology is converging in that new more memory efficient encoding algorithms are being developed and less expensive higher density memory chips are continually being introduced. The VISTA program is addressing voice response as an integrated complement to voice recognition. Presently a speech synthesizer is interfaced into AVRADA's computer facility for applications testing in the noise environment. The specific synthesizer used was selected for its ability to be programmed in-house. To this end a program was written (referred to as a Speech Editor) which enables various vocabularies to be stored on disk. These vocabularies can be accessed by several programs for intelligibility testing and applications testing in conjunction with the voice recognition equipment. Beginning in Phase III, various voice response units will be evaluated for intelligibility. Much of this intelligibility information should be available to the VISTA program through other Tri-Service efforts. The VISTA program will initially evaluate the available voice response intelligibility data for its application to the Army noise environment. Where data is still needed regarding the Army specific acoustical environment (i.e., radio and intercom systems as well as noise), the necessary testing will be performed under the VISTA program.

VISTA VOICE RECOGNIZER TESTING TECHNIQUES

At present the formulation of standards for recognizer testing is in its infancy and as yet no established recognizer testing standards or criteria exist. One of the chief difficulties in the formulation of standards for recognizer evaluation has been the determination of what criteria will yield meaningful information about the performance of voice recognizers in many diverse environments. This, unfortunately, creates the classic "chicken and the egg" problem for those who wish to apply this technology. The VISTA approach has been to devise a series of test and evaluation procedures applicable to the Army aircraft environment.

Figure 12 shows the typical test setup for recognizer testing. Recordings of aircraft noise are equalized and played into the sound absorption chamber. The output of a precision microphone, located in the chamber, is fed into a measuring amplifier and a spectrum analyzer to make adjustments

REED

in overall intensity and spectral content. The voice recognizer and a CRT terminal located in the chamber are connected to the Interactive Graphic Host Computer. Specialized software running on the host computer performs applications testing of the recognizers and recognizer comparative testing. Test results are then printed on the line printer.

The initial voice recognizer testing is limited to the two modes of recognizer operation--Unrestricted Vocabulary Search (UVS) and Restricted Vocabulary Search (RVS). UVS involves computer software which prompts the test subject (via CRT) with each word of a selected vocabulary. The computer permits the voice recognizer to search its entire vocabulary for a best match based on a predefined recognition threshold. If the utterance does not exceed the recognition threshold for any of the stored vocabulary words, a reject response is output from the recognizer to the computer. Likewise, if the voice recognizer perceives the utterance as noise, an appropriate response is output to the computer.

All responses generated by the recognizer are stored in a disk file. The RVS involves an Applications Simulation Program (ASP) running on the host computer. The ASP controls the voice recognizer as it would be in an actual aircraft application (i.e., radio control, navigation control, etc). At any point in the scenario the ASP restricts the voice recognizer to a specific subset of words in its vocabulary. Using radio control as an example, the ASP initializes to a command mode in which the user may request the status of a given radio. In command mode the recognizer is restricted to matching only those words which designate the various radios. At this point in the scenario it is neither necessary nor desirable to have the recognizer attempt to match an utterance to any other portion of the recognizer vocabulary. In actual field operation restriction of the vocabulary will decrease the occurrences of false matches by the recognizer and hence increase the reliability of the entire system. The RVS testing will give a measure of the relative reliability between restricted and unrestricted vocabularies as well as a closer measure of the performance which can be expected from a given recognizer in the field.

The following discussion describes a typical test session, delineating the test parameters and test results.

Radio control was selected as a candidate application because of the level of tasks to be performed and the fact that it is a non-flight control critical operation. Having selected radio control, the tasks to be performed were delineated and two basic functions were selected: the cycling of power and the selecting of frequency to a specific radio. The typical Army aircraft complement includes four radios: two VHF FM radios; one VHF AM radio and one UHF AM radio. Based on the above information, a vocabulary for the recognizer was devised (see Fig 13). It should be noted that the selected vocabulary contains two utterances which sound alike

REED

except for their endings ("Foxmike 1" and "Foxmike 2") and two utterances sound similar ("Victor" and "Enter"). These words were chosen to yield some information concerning the critical nature of sound alike words. Training patterns using the vocabulary of Figure 13 were made for several test subjects under four different conditions. The four conditions involve the use of two different microphones and training the recognizer with each microphone in the "Quiet" and in the "Candidate noise" environment. The noise environment selected for all training and testing is that of the UH-60 Black Hawk (the STAR testbed aircraft, see Fig 14). The training program resident on the host computer guides the test subject through the training process. After training is completed, the voice pattern stored in the recognizer is up-loaded into the host computer by a program called VOXDSK. VOXDSK handles all up-loading and down-loading of voice patterns.

Incorporated into VOXDSK is a mandatory request to the operator for header information concerning the test subject (Fig 15). The information includes the subject's name, the creation date and time of the voice pattern files, the condition of the test subject, the test conditions, the type of recognizer, the number of training passes and the number of words. VOXDSK reads the header file created by the operator and checks it for the required information. When VOXDSK is satisfied, it creates a composite file which includes the test subject header and the voice pattern (Fig 16). This self-documentation approach insures the traceability of voice pattern history and will minimize errors due to the incorrect use of voice pattern files. Having trained the recognizer, the testing procedures can now begin. This paper will be limited to a discussion of the procedures and preliminary results of the Unrestricted Vocabulary Search tests. The first step in the testing procedures is to create a test header file. This file is similar to the training header file except it refers to the specific test conditions which may be different from the training conditions. If a question mark is inserted into any information field (i.e., the "Date.Time" field) that data will be automatically requested from the operator at test time. Both the training and the test result files are uniquely numbered by a six-digit date field followed by a four-digit 24-hour time field separated by a period. When the test program runs, it creates a test results file. The operator is then prompted to enter the time of the test; the test program then looks up the voice pattern file used in the test and appends the training header to the test header in the test results file. For the UVS testing the subject is prompted, via the CRT, by the host computer with the vocabulary of Figure 13 stored in a prompt file. When the test subject responds to the prompt, the recognizer outputs its best match response (or no match response, if below the rejection threshold or noise) to the host computer. The host computer outputs all recognizer responses to the test results file for later comparison. When the data in the prompt file is exhausted, the testing program terminates. A comparison program automatically initiates at the termination of the test program and outputs all the header information contained in the test results file to

REED

the line printer; the program then begins comparing the response data in the test results file to the original prompt file. The results of the comparison are output to the line printer along with three columns of information (Fig 16). The information includes: an accuracy column with the number and percentage of totally correct responses; a reliability column which contains the number of totally correct responses plus the number of negative responses (a negative response is one in which either the recognition threshold is not exceeded or the recognizer perceives noise) and the percentage of same; and a latency column which is the difference information between reliability and accuracy. Because Reliability combines the total number of correct responses with the total number of negative responses it yields the percentage of the time (based on usage) the recognizer will not get the aircraft into trouble by going into the wrong mode of operation. The latency data yields the percentage of time a response would have to be repeated.

For the radio control vocabulary, an entire test run can be completed in an average of one minute and thirty-five seconds. For the same test conditions ten contiguous runs are performed by the computer in a total time of approximately 16 minutes. It is evident that, due to the rapidity of the testing procedures, many test condition variations can be tried in a relatively short period of time.

TEST RESULTS

The following discussion will be limited to the preliminary tests which were performed using the UVS technique. The results of those tests are summarized in Figures 17 and 18. The following conditions were adhered to for all testing: UH-60 noise at 103 dBA (Fig 17) and 107 dBA (Fig 18); microphone position just brushing the test subject's lips; the recognizer connected to the standard Army aircraft intercom system; the vocabulary of Figure 13 trained with five passes per word; approximately one second delay between word prompts; and a recognition threshold of 105 (NOTE: Each pattern is composed of 128 bits; a recognition threshold of 105 would require the recognizer to match 105 of 128 bits before outputting a match response).

The test results of Figures 17 and 18 represent the compilation of up to eighty tests per subject (10 iterations per test condition). They are ordered from the highest to the lowest accuracy for each test subject. It can be immediately observed that for each test subject, the highest recognizer accuracy was achieved for the vocabulary trained in the actual noise environment. This confirms the results of similar tests performed in different environments by other agencies. Another important feature of the test results is that in all cases, the Electret microphone achieves the most accurate results when the vocabulary is trained in the noise and the least accurate results when training in the noise is not employed. This

REED

apparent paradox can be explained by comparing the frequency response characteristics of each microphone (Fig 19). The Electret microphone has a broader frequency response; therefore, the contribution of noise to the pattern generated by the Electret microphone is more significant than the same pattern made in the noise with the M-87 microphone. Although the wider frequency response of the Electret microphone passes more noise, it also passes a greater portion of the low-frequency spectrum. This appears to account for the Electret microphone producing the most accurate recognition results for every test subject. The apparent ability to trace recognizer performance to microphone characteristics does provide a means of test result validation. As the number of test subjects increases, particular attention will be given to see if this trend continues.

Even now, a deviation from this trend is used to prompt a reexamination of a test subject's training pattern for accuracy.

RESEARCH AREAS

In the near-term testing of the VISTA program, we will be concentrating on the following areas: the front-ending of recognizers with off-the-shelf noise cancellation devices; the significance of Army aircraft vibration on human speech; the establishment of a restricted vocabulary for near-term application; the electrically vs acoustically mixing of noise and speech for training purposes; and the effect noise has on the character of human speech. The latter research area will be a joint effort by the Army (AVRADA, Ft Monmouth) and the Navy.

To date, many different techniques have been developed for the cancellation of noise. Many of these techniques have resulted in a significant loss in intelligibility; however, this loss of intelligibility, while significant to a human listener, may not affect voice recognition equipment. The VISTA program will therefore investigate the effects of applying existing noise cancellation techniques to voice recognizers.

The use of a reference voice will play a significant role in the investigation of vibration on human speech. The reference voice will be created using a recording of a selected subject played back through an artificial voice transducer. This reference voice will be used for training as well as testing the recognizer. Recordings will be made of test subjects and the reference voice in both the chamber aircraft noise environment and the actual aircraft noise environment. Because in actual flight the reference voice will not be subject to vibration, it is hoped that a comparison of the various test results will yield meaningful and repeatable information regarding the effects of vibration.

REED

AVRADA will maintain a complete electrical hot bench of the STAR aircraft. This will provide an ideal environment for meaningful human factors work directed toward the implementation of voice recognition equipment in the Army aircraft environment. It is AVRADA's desire to enter into a cooperative arrangement with other governmental human factors agencies whereby those agencies would utilize AVRADA's STAR hot bench and STAR aircraft to perform human factors analyses in a relevant Army aircraft environment. Through this human factors work, a restricted vocabulary for near-term applications will be defined.

Because of the high ambient noise found in Army aircraft, it is undesirable to require training in the noise. Therefore, experiments will be conducted in electrically mixing the noise with the speech in the quiet for recognizer training. The resulting test data will be compared to the same test conditions using training patterns made by acoustically mixing the speech with the noise at the microphone. Because the speech in the former case will be generated in the quiet, as far as the test subject is concerned, any effect the noise has on the speech will not be reflected in the electrically mixed voice pattern. To investigate the effect of noise on speech, experiments will be conducted by subjecting the test subjects to noise via a high-quality headset. Care will be taken to insure that the test subject hears the same level and balance of sidetone (subject's own voice) and noise in the headset. The subject's voice, which is now essentially in the quiet (because the only noise is in the headset which has minimal leakage), will be recorded and analyzed for aberrations traceable to the effect of the noise. The intent of the speech analysis is to determine if an apparatus can be devised to artificially shape speech produced in the quiet, giving it the characteristics of speech produced in the noise.


CONCLUSION

➤ Although the preliminary test results are encouraging, it must be remembered that they were taken under ideal conditions. For all testing, the microphone was positioned just brushing the test subject's lips; however, a test was run with one test subject placing the microphone approximately four millimeters from the test subject's lips. The test results showed a 50% decrease in recognition accuracy for the same conditions as those with a microphone touching lips. Although the results are preliminary, it is apparent that the signal-to-noise ratio is a key factor in recognition accuracy. Another problem arises because of the automatic gain controls (AGC) found in most aircraft intercom systems. When there is no voicing for a period of time, the AGC increases the intercom sensitivity. If the first utterance spoken is intended for the recognizer it will likely be rejected because of the distortion caused by the AGC adjusting the gain during the utterance. This is demonstrated in the test results of all test subjects. No attempt was made to set the AGC

↓
cont

REED

Cont

before beginning the test; as a result, 90% of the first utterances were rejected which resulted in the lowering of the accuracy score by approximately 4%. The AGC has a release time of 10 seconds and the prompts are issued every second; therefore, after the first utterance the AGC has little effect. Some side tests were performed by making an utterance before signaling the computer to begin the test, and in each case the accuracy of the first test word increased to a point comparable to the other vocabulary words. 

The VISTA program is the first in-depth attempt to apply voice recognition and response to the Army aircraft environment. Participation by other governmental agencies is being sought for cooperative efforts utilizing AVRADA facilities for the application of this technology to the Army aircraft environment.

END
DATE
FILMED
8488

NT

AD-A117 288

VOICE INTERACTIVE SYSTEMS TECHNOLOGY AVIONICS (VISTA)
PROGRAM(U) ARMY AVIONICS RESEARCH AND DEVELOPMENT
ACTIVITY FORT MONMOUTH NJ L W REED 18 JUN 82

2/2

UNCLASSIFIED

F/G 17/2

NL



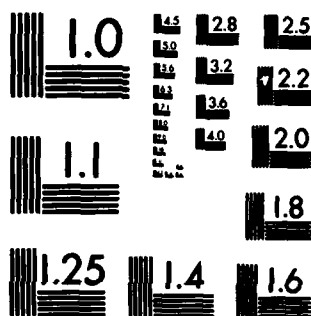
END

DATE

FORMED

583

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

SUPPLEMENTARY

INFORMATION

AD-A117 288

ERRATA

The 19 figures referenced in the paper are not available per the author.

DTIC-DDAC
6 May 83